UNCLASSIFIED

AD NUMBER

AD213781

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to U.S. Gov't. agencies and their contractors; Administrative/Operational use; 8 Dec 1958. Other requests shall be referred to Rome Air Development Center, Griffiss AFB NY 13441-5700.

AUTHORITY

Per OTS ltr, 2 Jun 1967

THIS PAGE IS UNCLASSIFIED

UNCLASSIFIED

AD 213 781

Reproduced
by

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

RADC-TN-59-136

AD-213781

DECEMBER 8, 1958

THE RELATIONSHIP OF INFORMATION-USE STUDIES AND THE DESIGN OF INFORMATION STORAGE AND RETRIEVAL SYSTEMS

SAUL HERNER

HERNER AND COMPANY
906 NEW HAMPSHIRE AVENUE, N. W.
WASHINGTON 7, D. C.

TECHNICAL NOTE NUMBER 1

CONTRACT AF 30(602)-1857

PROJECT NO. 4594

TASK NO. 45940

PREPARED
FOR
ROME AIR DEVELOPMENT CENTER
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
GRIFFISS AIR FORCE BASE
NEW YORK

Table of Contents

Abstract

~~The purpose of this~~ report is to review the results of past studies of the information-gathering methods of workers in various fields, and ~~to demonstrate~~ the general applicability of these results to the design and improvement of information programs and systems. Two cases from the literature are used to illustrate storage and retrieval systems that do and do not meet the requirements of their users, and to show the contrasting need of the pure scientist for mere references to information ~~on the~~ one hand and of the applied scientist for direct access to actual information. ~~on the other.~~ The applied scientist is shown to require the services of large storage and retrieval programs manned by highly trained personnel, while the pure scientist is best served by a conventional library in which the publications are arranged on the basis of a classification scheme ~~that is~~ reflective of the scientist's customary association of subjects. ~~General classi~~fications are rejected because they are too broad and redundant, and special classifications based on the ~~habits and requirements of the user bodies are recommended and described.~~ Low redundancy in ~~both~~ manual and machine systems is cited as a means of improving input and output.

The logical requirements of typical questions put to retrieval systems by users are shown to be simpler than those for which most systems are being designed. Differences in the required information coverage of systems serving pure scientists, applied scientists, and social scientists and humanists are indicated, with the pure scientist requiring a narrow collection going back comparatively far in time, the applied scientist requiring a broad but shallow coverage, and the social scientist requiring both depth and breadth. Investigations of the various media used by user groups to communicate and obtain different kinds of information are advanced as a further means of knowing what should and should not be put into a storage and retrieval system. The value of past user studies as first approximations, and the need for further studies, are cited.

Introduction

Perhaps the most important and least considered factor in the design of information storage and retrieval systems is the user of such systems. Regardless of what other parameters are considered in the development of a storage and retrieval mechanism, it is necessary to consider its potential use and mode of use by the persons or groups for whom it is intended; it is necessary either to fashion the system to suit the user's needs, habits, and preferences, or to fashion the user to meet the needs, habits, and preferences of the system. Both approaches are possible, but the second one, involving the education and re-education of the user, is evolutionary and futuristic. A system designed for *now* should at least be able to serve the present user.

This is not to say that the present user will not change as his experiences and opportunities in getting information broaden; he will inevitably change. One striking feature of most successful storage and retrieval systems is that they start out to meet current needs of specific user groups, and, gradually, as the systems meet more and more challenges put to them by the users, they themselves are forced to change because the demands placed upon them begin to exceed their capabilities. But these successful systems begin as means of meeting actual, immediate user requirements, on the users' terms and not on those of the system.

One thing that has seriously hampered the development and use of advanced information storage and retrieval systems has been the tendency of the designers to think of the user as a constant. Frequently, the problem of defining this constant is solved by the designer, who assigns himself the role of typical user and extrapolates from his own information requirements and experience in developing his system. The usual result is that the system suits the designer's needs, but not those for whom it is designed.

As is true of most utilitarian items, it is very difficult to design effective information systems in the abstract. Before entering into the engineering and design phases of a system, it is necessary to know just what the system is supposed to do and for whom. This information can only be obtained by studying the habits and requirements of the potential user.

There have been innumerable user studies performed over the past decade or more. These studies, which have utilized a wide variety of analytical methods, have been reviewed in papers by Egan and Henkle (3), Shaw (13), and Tornudd (16). Unfortunately, these reviews and most recent discussions of user requirements have been more concerned with method than with applicable results. In their preoccupation with <u>how</u> past studies were done, the reviewers and discussants have overlooked the fact that most of them were done by working librarians and information specialists for specific practical purposes connected with the improvement of existing information programs. While admittedly imperfect in conception and execution, the majority of these studies have produced results, in the sense that they have furnished operators of information programs insight as to how these programs are used or are likely to be used. A striking characteristic of these results is that despite the fact that they have been derived by a variety of means they have corroborated one another in a number of important respects. The purpose of this report is to review some of the more consistent results of past studies, and to demonstrate the general applicability of data on user habits and preferences to the design and improvement of information programs and systems.

Illustrations of User- and Non-User Oriented Systems

The literature is replete with descriptions of information programs and systems that meet actual, present user requirements and of those that do not. Among those that are designed around the actual needs of the user group, the method used to determine these needs varies with the organizational activity, the size of the user group, organizational policy, and other factors. In some cases, the method consists of group discussions between the designers and users of the program or system; sometimes a simple questionnaire is used; in other cases, face-to-face interviews are used; in still other cases, the analysis is done by examining circulation files and other records of literature use. But, regardless of the method used, user-oriented systems have in common the characteristic of being tailored for a specific group, and of being designed around the preferences, habits, and foibles of the group. They are, in short, designed to meet the user's needs, on the user's own terms.

The Smith Kline and French System

One very interesting example of a system designed to meet a specific, carefully-defined set of user requirements is that of the Smith Kline and French Laboratories. The function of the system, which is described in part in a paper by Rockwell, Hayne, and Garfield (12), is to store, search, and correlate pharmacological and clinical information on various drugs of interest to the company. The system utilizes an IBM 101 Statistical Machine and an indexing code reflecting the various aspects of the drugs that are likely to be of interest to the searcher. Originally, information on pertinent drugs was searched by means of a conventional card index. However, it became evident at a relatively early stage in the evolution of the system that the card index did not offer the depth of document analysis and the opportunity for data and concept correlations which were necessary for the full exploitation of the collected information. This led to the development of a mechanized system.

In terms of coded documents, the system is relatively small, totalling 30,000 items, the majority of which are unpublished reports of clinical and laboratory data. However, in terms of intensity of indexing of the information contained in these items, the collection is very large. In terms of what can be done with the available information, by way of detailed searches and correlations of diverse facts and data, the system is extremely sophisticated. But the thing that makes it truly sophisticated is the fact that it is designed around the clearly defined needs and interests of its users.

To illustrate this point, the strictly pharmacological information in the system--that having to do with such fundamental things as the mode of action of drugs--is indexed in such a way as to produce bibliographies rather than actual data. The reason for this is that the pharmacologists in the company were found to prefer to do their own reading, correlation, and synthesis, and all they want is references to the pertinent literature; they do the rest. On the other hand, the clinicians in the company--those concerned with the action of drugs on human patients--prefer to receive actual data, and, if possible, they want it correlated and tabulated for them. Therefore, clinical information is entered into the system in such a way as to permit routine correlation and tabulation.

This difference in the approach of pure and applied scientists or in this case pre-clinical and clinical scientists to information is one which has also been found in other user studies (5,6). An important dimension that Smith

Kline and French has added is the actual application of the results of its own and other studies to the improvement of a working system. By applying these results, the firm has been able to turn a relatively unpretentious system into one of great effectiveness, and has ensured that it is used to the total extent of its capabilities.

One other important attribute of the Smith Kline and French system is its dynamicism. The designers of the system are keenly sensitive to changes in user-requirements: changes arising from shifting subject interests and from a growing awareness of the user-group as to the capabilities of the system. By keeping aware of shifting search demands and interests, the designers of the system are able to ensure that it meets the real needs of the group for which it is designed and thus retains its vitality.

The NOTS System

In striking contrast with the Smith Kline and French system is one developed at the Naval Ordnance Test Station (NOTS), and described in papers by Burtnett (2) and Bracken and Tillitt (1). The NOTS system uses an IBM 701 Calculator to store and search coded Uniterm entries for a collection of approximately 20,000 documents. The average document is indexed under eight entries.

According to Burtnett, the machine, using magnetic tape as a storage medium, will perform the average search in about one minute, while it takes a human operator eight minutes to do the same search manually. Burtnett gives the average cost of searching with the 701 as $2.20 per minute. As of the time that the paper by Bracken and Tillitt was written (1956), the machine was being used for an average of 48 searches a week, requiring 48 minutes, as opposed to the six and a fraction manhours it would take to do the searches by hand.

With 48 searches a week, the manhour savings derived from the use of the 701 are quite small, amounting only to slightly more than five hours a week. But the cost of these manhour savings is very high. Assuming that a human searcher is paid $5.00 an hour (probably far too high a figure), the total cost of doing the 48 searches by hand would be slightly more than $30.00, while the cost of doing the searches on the 701 would be $105.60. This high cost is inevitable when the distance between what a machine can do and what it is called upon to do is too great. Actually, the machine is being given all the advantages in the foregoing comparison. If the cost of putting the information into the machine and the cost of programming to get it out again were taken into account, the average search on the 701 would come to considerably more than $2.20. It would probably come closer to $10.00.

But all this is incidental to the fact that the NOTS system is not really designed to do the things that its users require it to do. For the most part, ordnance testing is concerned with applied science and engineering. As indicated earlier, applied scientists, either as a result of their training, or possibly due to the tight time schedules under which they are generally forced to work, prefer to get their information in finished form; they generally want immediately-usable information, rather than references to documents containing it. The NOTS system is patently unsuited to such a requirement. It delivers references only, and, to make matters worse, these references come in the form of document accession numbers rather than meaningful bibliographic citations. Under such a circumstance, the person handling a reference request has either to translate the accession numbers into citations, if a bibliography is required, or he must deliver the documents themselves, forcing the requester to wade through a good deal of relevant and irrelevant material to find what he needs. Either alternative adds materially to the time and cost of the search, and yet neither fully satisfies the needs of the particular user body for which the system is intended.

Systems for Pure vs. Applied Scientists

The experience of the Smith Kline and French Laboratories in the development of a dichotomous retrieval system to provide for the diverse information requirements of its pre-clinical and clinical scientists is illustrative of one generalization which can be drawn from user studies. The design of the Smith Kline and French system confirms a hypothesis evolved from interview studies of the information-gathering habits of scientists in academic and industrial settings (5,6). This hypothesis is as follows: Information storage and retrieval systems serving pure scientists are best designed as two-stage operations, in which the user is furnished bibliographic references and does his own selecting, reading and interpreting of items in the bibliography, while the applied scientist is best served by the one-stage type in which the user is given actual answers to questions, rather than references to literature containing the answers.

The Applied Scientist

The foregoing hypothesis, simple as it is, gives rise to a number of broad implications relating to the design of storage and retrieval systems. One implication is that the intellectual demands on the operator of a system serving pure scientists are significantly lighter than those placed on the operator of a system serving applied scientists. In the first instance, all that is required of the system is a list of references, and nothing more. In the second instance, specific answers to specific questions are required. Producing actual answers to questions implies a far more complicated input and output mechanism than does the production of mere references. When answers rather than references are required, it is necessary for the system operator to have selected, read, interpreted, and encoded potentially useful information in such a way that it can be retrieved in a form that is immediately applicable to user problems or inquiries. This means that the person who puts the information into the system must be at least as conversant with the subject matter dealt with as the user for which the system is designed; it means that the coding parameters must be so finely resolved as to permit the assimilation of highly specific bits of data and other information, in a way that permits ready retrieval in an intelligible form; it means a highly dense memory; and, finally, if this memory is to be exploited profitably, it means an output mechanism that is capable of searching and correlating data and other types of information at high rates of speed.

The Pure Scientist

By contrast, a two-stage system serving pure scientists has a much easier job to do. The reason for this relative simplicity is a negative one: owing to the high level of sophistication and the narrow fields of specialization in pure science, it is extremely difficult for anyone but the requester himself to perform the selection and interpretation phases of a search. Literature selection and interpretation, and the subsequent process of synthesis, constitute, for the pure scientist, the very essence of creativity. This is a hard thing to have done by proxy.

Thus, the person putting information into a storage and retrieval system serving pure scientists, must, more or less by default, do his encoding in only the most general terms; he must describe whole published units, such as books, papers, articles, etc., rather than the specific information contained in these units. This means a storage and retrieval system in which the primary job of the operator is not to exploit a collection of information on behalf of a user group but to arrange the collection in such a way that the user can find what he wants by himself. This logically connotes a conventional library in which publications are arranged according to their subject content.

The Retrieval Capabilities of Systems

The problem of helping the user to "find what he wants by himself" poses an important challenge for the library serving the pure scientist. Two recurring findings of past user studies (5,6,14) are, first, that pure scientists prefer to come to the library in person to consult the literature, and, second, that they do not use the library card catalog, preferring instead to go directly to the shelves and browse. If the library serving a body of pure scientists is to promote the best possible use of its collections, it should arrange them in such a way as to reflect the logic and preferences of the specific group that is to use them.

General Library Classification

Most special libraries approach the problem of shelf arrangement by adopting one of the major classification systems, such as the Library of Congress Classification, the Dewey Decimal Classification, or the Universal Decimal Classification. The major library classification systems share the common shortcoming of most storage and retrieval systems: they treat the user as a constant; their structure and design are based on the content of the literature, rather than on the way that the literature is used. As a result, the literature in a given special library is likely to be arranged in the exact same way as it is in a majority of libraries, regardless of the specific interests or viewpoints of the clientele of the given library. This is actually no more realistic than using the same type of facilities to house organic chemists and theoretical physicists.

Classifications such as Dewey Decimal, Library of Congress, and Universal Decimal, share a failing that is common to all systems that attempt to be universal in outlook: they try to see everyone's viewpoint; they try to encompass all the world's literature and to classify it as every possible user in every possible clime would classify it. This makes a very difficult job for the cataloger whose job it is to find the right niche for each piece of literature he receives, and for the searcher who must find out what the cataloger did with it.

User-Based Coding Schemes

There are, on the other hand, a number of classification schemes that have more definite viewpoints. One basic characteristic of these schemes is that they are designed for specific user groups, rather than for all literature readers. Another basic characteristic of such classification schemes is that they treat the problem of organizing a literature collection as a purely physical one, in which the goal is simply to categorize pieces of literature in such a way that the user can easily find them. The question of how a given user group habitually defines subject classes or categories of literature is generally related to the purposes for which the literature is used. One method that has been employed to define purpose of use has been the analysis of typical reference requests made by members of the user group over a representative period of time (8). From this, a set of typical search parameters or requirements are developed, and the literature is categorized under them.

Besides the obvious advantage of permitting a collection of documents to be laid out in a way that the user is likely to look for them, classifications keyed to specific user groups, and based on a detailed knowledge of the technical habits and requirements of these groups, permit important improvements in input. The improved input in user-based classification schemes stems from the fact they minimize redundancy by organizing a given literature collection from only one viewpoint, that of the group that is to use it. This lack of redundancy greatly simplifies the work of the classifier by clearly delimiting the areas in the collection into which a given piece of literature can go.

In addition to speeding up the classifying and searching processes in conventional, open-shelf libraries, user-based classifications offer the same input and output advantages for machine systems in which classifications are used as the basis for storage and retrieval codes. Anything that can diminish the amount of time it takes a coder or cataloger to decide the subject categories covered by a given document, and anything that will simplify the decisions of the programmer in deciding how to get the information out of the system, will obviously improve the economy of a machine system.

The advantages of user-oriented systems are not limited to those with classification schemes as their bases. There are a variety of cases in which user-orientation has contributed to the success of non-hierarchical indexing and coding systems. Mooers (11) has described what he terms a "Descriptor Dictionary System," in which the indexing and code terms are developed as a special vocabulary reflecting the requirements and viewpoints of the specific group that is to use the system. By treating each indexing and coding vocabulary as a special case, Mooers is able to limit severely the number of terms or descriptors, and to produce a corresponding speed-up in input and output.

Another example of the diminution of an indexing vocabulary by making it coincide with the most likely search parameters of the user-group is described by Wall (19), who applied this method to the design of a "peek-a-boo" index to pressure vessel drawings. By consulting with potential users of his index, Wall found that he was able to reduce his indexing or searching parameters from 98 to

20. These 98 parameters that he started off with were those that the pressure vessel drawings in his collection might be searched by. The 20 that remained at the end were those by which they actually were searched. This diminution of indexing and search parameters, and the exclusion of those that are unrealistic, obviously speeds up the input and output processes.

There is yet another advantage in designing an indexing system around the pre-determined preferences and requirements of a given user-group, and in the resulting diminution in the number of indexing and searching parameters. This additional advantage has to do with machine searching systems. One of the factors limiting the speed of machine sorting and scanning systems is the number of characters used to represent each subject. As a rule, the greater the number of characters per subject, the slower the search. This has particular pertinence to systems utilizing digital computers, in which each subject code is scanned character by character. But it also applies to punched cards, in which the number of subjects that can be punched on a card is dependent on the number of characters used to represent each subject. With a diminution in the number of subjects in a system, it becomes possible to use a shorter and simpler notation to represent each subject, and there is a subsequent improvement in sorting or scanning efficiency.

The Logical Capabilities of Systems

Another phase of the searching mechanics of a system that can benefit from a knowledge of the habits and requirements of its users is concerned with the ability of the system to perform searches involving correlations between two or more subjects. Obviously, if the questions put to a system habitually involve correlations which exceed its capabilities, the system is ineffective.

Likewise, if the correlation capabilities of the system exceed the complexity of the questions customarily put to it, there is a resulting waste.

The question of the kind of search demands that requesters are likely to make of storage and retrieval systems has been investigated in two recent studies (7,20). Both of these studies involved the analysis of a representative sample of reference questions produced by a specifically-defined body of users. In one case, the body of users consisted of members of the scientific staff of the Tonawanda Research Laboratory of the Linde Company, while in the other it consisted of the personnel of 14 national laboratories and prime contractors of the Atomic Energy Commission.

As one might imagine, the detailed results of the two studies varied, since the makeup of the two user-groups were quite different. However, in one very basic respect there was agreement. In both studies, in which actual examples of search questions were analyzed on a statistically valid basis, it was found that the questions produced by the groups investigated were considerably less complicated than the questions that the majority of retrieval systems are designed to handle. They generally covered fewer concepts, and required far simpler logical manipulations than would seem likely from most discussions of storage and retrieval systems in the documentation literature. This may of course be due to the fact that requesters limit their searches to the capabilities of the retrieval facilities available to them. But it is also undoubtedly due, in part at least, to the failure of systems designers to take the actual requirements of the users of these systems into account.

Content and Scope of Systems

Reference has already been made to the need for higher density memories and speedier and more sophisticated searching mechanisms in systems serving applied scientists than is necessary in those serving pure scientists. This need was ascribed to the fact that applied scientists generally want actual answers to search questions, rather than references to documents containing these answers, and this means more intensive analysis and finer indexing of the information collection. There is still another reason for the need for more versatile systems for groups consisting of applied scientists. This is related to the fact that applied science, as a rule, crosses more subject areas than pure science, and it is therefore more difficult to encompass the sources of information required by the applied scientist in a single storage and retrieval system.

Subject Coverage

The question of "subject dispersion," or the use of the literature in one subject by workers in another, has been dealt with by a large number of investigators who have, among them, covered most fields of pure and applied science as well as the social sciences and humanities. The method used by the majority of these investigators is called "reference counting." Briefly, "reference counting" consists of statistical analyses of the types and quantities of publications referred to by authors in representative publications in a given field. Stevens (15) has summarized and synthesized various of these studies, and has arrived at a number of conclusions regarding not only "subject dispersion" but also the temporal span of usefulness of the literature in the various fields studied and the comparative number of different publications that workers in a given field have to consult in order to get the information they need. Needless to say, these two additional factors can also have a profound effect upon the design of a storage and retrieval system.

Regarding "subject dispersion," Stevens found, first of all, that the applied scientist is far more likely to use the literatures of fields outside of his own than the pure scientist. Second, he found that the worker in a science that is relatively new is likely to make greater use of literature outside of his field than the worker in an older field. Thus, in fields such as mathematics, chemistry, and physics, there is a much smaller "subject dispersion" than in such fields as biophysics and biochemistry, which are still too young to be completely independent of the fields from which they stemmed. Both Stevens, and Voigt (18), who did a "reference counting" study covering the literatures of agriculture and engineering, have demonstrated that both the pure and applied scientists adhere to the rule that the use of literatures outside the immediate field varies inversely with the age of the field.

As for the question of "title dispersion"--the number of publications that researchers have to consult to meet their information requirements--Stevens' findings are essentially the same as in the case of "subject dispersion," with a greater scatter of titles consulted among applied than pure fields, and among new sciences as opposed to old sciences.

Departing from the natural sciences and technologies, and analyzing the results of "reference counting" studies in the social sciences and humanities, Stevens concluded that workers in the non-scientific subjects exhibit far more scatter than pure and applied scientists, in both the subject literatures they consult and in the number of different publications they use to obtain the information they require.

Time Coverage

The useful life span of the literature in a storage and retrieval system is still another factor governing its size and design. Measurements of the comparative ages of the literature used by workers in various fields have been made in a number of studies. These studies have utilized several different techniques, including "reference counting" (4), face-to-face interviews (5), and analyses of library circulation records (17). While the results of these studies have varied as to particulars, they have produced striking agreement on certain basic points.

One major point of agreement regarding the temporal span of use of scientific and technical literature concerns the pure and applied scientists once again. As a rule, pure scientists go considerably farther back in their use of the literature than applied scientists. Several reasons for this difference have been advanced. Primary among them is the fact that the pure scientist is by nature a scholar; he works at a leisurely pace and has time to be more thorough in his use of the literature than the applied scientist, who is forced to sacrifice scholarship for expedience and to be somewhat more shallow in his reading than the pure scientist. This explanation is borne out by the finding that the applied scientist makes less use of written and published information in general than the pure scientist, leaning more on personal contacts for his information (5). It is also borne out in the applied scientist's greater willingness to delegate his literature searches to other persons or agencies.

Another likely reason for the use of more current information by applied scientists is the fact that applied science is far more changeable than pure science. The ideas and techniques of pure science apparently retain their usefulness and timeliness longer than do those of applied science. The relatively rapid turnover of ideas and trends in applied science gives rise to a rapid obsolescence in its literature.

This difference in dynamicism and the subsequent differences in the rate of literature obsolescence apply not only to pure and applied science as a whole but to specific fields of science as well. For instance, it has been shown that physicists use literature of much more recent origin than chemists, who tend to go back farther in time (15). Chemists, on the other hand, do not go back as far in the literature as biologists (10). Comparing natural science and technology with the social sciences and humanities, it develops that the social sciences make use of still older literature than natural science and technology (15). Thus, we have a further indication that some fields are more dependent on change than others, and that this changeability is reflected in the literature of the fields involved.

Focusing on science and technology in particular, it is interesting how relatively recent in origin the bulk of the literature in current use turns out to be. Past studies indicate that for most fields the outside age of the literature consulted is 15 years. However, it is important to be conservative in applying this figure in deciding what goes into a storage and retrieval system, since workers in some fields typically depend on literature which is far more current, while workers in other fields go back considerably farther than 15 years. But it is evident that regardless of field the temporal span of the literature used is completely measurable and finite, and can be used as a partial guide in the design of storage and retrieval systems for specific user groups.

Types of Sources of Information

One other extremely important consideration in the definition of the content and scope of a system is the derivation of the information that goes into it. Sources of information vary from field to field and from activity to activity. In the applied phases of physics and chemistry, for instance, a very high proportion of the useful written information comes from unpublished progress reports (5). This would seem to relate to the dependence of applied scientists on information of an extremely current nature. By contrast, physicists and chemists involved in basic or pure research lean primarily on research periodicals for their written information. On the other hand, biologists make greater use of textbooks and treatises than either chemists or physicists, this apparently stemming from the fact that biology is not subject to as rapid a rate of change as chemistry and physics (10). Thus, it is extremely difficult to predict the kinds of literature that should be covered by a system without a careful study of the people who are going to use it.

In addition to telling the system designer what should go into a system, a study of its potential users can also tell him what should not be included in it. Hertz and Rubenstein (9) have demonstrated that certain types of information are generally communicated one way, while other types are communicated in other ways. Some kinds of information are almost always communicated orally, others are communicated by informal written means such as memoranda, still others are communicated by means of formal vehicles such as research reports, periodicals, and treatises, etc. The specific vehicles used for different kinds of information vary in turn with the activities of the individuals and groups involved. The results of the Hertz and Rubenstein study are indicative of the fallacy of attempting to design and operate a storage and retrieval system without an intimate knowledge of the information-gathering and communication habits of the individuals and groups that are to use it.

Conclusions

The primary conclusion that would seem justified from an analysis of user studies done to date is that, imperfect as they are, they can still be an important source of guidance in the design of storage and retrieval systems. Even among the studies discussed in the present report, which constitute but a small fraction of those that have been done, there are beginning bases for useful generalizations on the handling and organization of information for retrieval.

There is, for instance, a strong indication in the findings of the studies discussed that pure scientists are not likely targets for the efforts of documentalists in the design of ornate centralized information storage and retrieval programs, first, because the relatively concentrated nature of their literature negates the need for such programs, and, second, because the nature of their work makes it extremely unlikely that they would delegate their information chores to outside agencies.

There is, however, evidence that applied scientists can be helped by large centralized storage and retrieval mechanisms. Applied scientists already customarily delegate their searches when they can, making far more use of reference services than pure scientists. This is apparently related to the fact that the use of the literature does not constitute as intimate a part of the research process as it does for pure scientists. In addition, the degree of scatter in the sources of applied research information makes it impractical to break these sources into smaller units covering specific subject areas, as is possible in the case of pure science.

On the other hand, the relatively ephemeral nature of applied research information creates a problem in its organization. In order to attain maximum usefulness, applied research information collections must apparently be organized in such a way as to permit direct answers to questions rather than references to sources of answers. However, this necessitates highly detailed content analysis on the input side and a very speedy and sophisticated searching mechanism on the output, which, in turn, raises the question of whether it is worth it. The answer to this question will vary from situation to situation, depending on just how important the information is to its users, and what alternatives they have in obtaining it. Here again, the advantages of knowing the habits and requirements of the user group commend themselves.

The social sciences and humanities pose still another storage and retrieval problem. Social scientists and humanists go back farther in their use of the literature than even the pure scientists, and yet the kinds of literature they use are far more diverse than those of even the applied scientist. To make matters worse, the social scientist and humanist place great reliance on the literature. Indeed, for many and possibly most, the library is as crucial as the laboratory is to the physical and biological scientist.

A certain proportion of the solution of the information problem of the social scientist and humanist is automatically derived from the fact that they are extremely heavy users of their literatures, and that the nature of their work, and the training that precedes it, make it easier for them to fend for themselves than it is for the natural scientist, and particularly the applied scientist whose focus is less on the library and more on the laboratory. Nevertheless, there are undoubtedly areas in the literature of the social sciences and humanities that lend themselves to change in methods of storage and retrieval. More exact data on just how scientists and humanists presently meet their information requirements, insofar as they are able to, would perhaps reveal what areas are subject to improvement and how this improvement might be brought about.

Returning to the original theme of these conclusions, probably the most interesting aspect of the foregoing discussion is that it would not have been possible were it not for the fact that a good deal of meaningful evidence concerning the information-gathering habits of users of information is in existence now. More such evidence is needed, but more use has to be made of the evidence that is already at hand. It is probable that the intelligent use of this evidence would result in solutions or partial solutions to many problems which have been subjected to general and perennial theoretical discussions. The least it could do is tell us what we still do not know, and what studies of user habits and requirements remain to be done.

It may be argued that the evidence which has been produced by past studies, and which will probably be produced by future studies, is of a crude or approximate nature. However, even a crude measurement is better than none at all. Most analytical undertakings start with first approximations which are refined on the basis of subsequent tests and experience. But before there can be refinements there has to be something to refine.

Science itself is continually changing as newer and better methods of analysis and measurement are evolved. The scientist, knowing this, might sit and wait for the ultimate method to be developed. But of course he doesn't; he makes do with the tools at hand, and he advances the cause of knowledge as best he can. This approach commends itself to the documentalist.

References

1. Bracken, R. H., and Tillitt, H. E., Journal of the [illegible] Machinery, 4, 131-6 (1957).

2. Burtnett, I. K., in "Conference on Multiple Aspects [illegible] Retrieval," pp. 83-90, Washington, D. C., Armed [illegible] mation Agency, 1957.

3. Egan, M., and Henkle, H. H., in Shera, J. H., Kent [illegible] eds., "Documentation in Action," pp. 137-59, New [illegible] ing Corp., 1956.

4. Fussler, H. H., Library Quarterly, [illegible], 19-[illegible]

5. Herner, S., Ind. Eng. Chem., 46, [illegible] (1954).

6. Herner, S., Chem. Eng. News, 34, [illegible] (19[illegible]).

7. Herner, S., and Herner, M., Area [illegible] preprints of [illegible] Conference on Scientific Information, 1958, [illegible] Academy of Sciences, 1958.

8. Herner, S., and Meyer, R. S., [illegible], 125, [illegible]

9. Hertz, D. B., and Rubenstein, A. [illegible] Team [illegible] Technical Publications, 195[illegible]

10. Hintz, C. W. E., "International [illegible] Scholarship [illegible] University of Chicago, 195[illegible]

11. Mooers, C. N., Aslib Proceedings [illegible] ([illegible]).

12. Rockwell, H. E., Hayne, R. L., and [illegible], [illegible] 726-31 (1957).

13. Shaw, R. R., in "Pilot Study on the Use of Scientific [illegible] tists," Washington, D. C., National Science [illegible]

14. Stevens, [illegible] E., "Summary of the Literature on the [illegible] Worker of the University Library Catalog," Uni[illegible] School Occasional Papers No. [illegible], Urbana, [illegible] brary School, 1956.

15. Stevens, [illegible] "Characteristics [illegible] Subject Literature [illegible] [illegible], pp. [illegible]-21, Chicago, Association of College [illegible] 1953.

16. [illegible], E., Area I, Preprints of Papers for [illegible] Scientific Information, 1958, pp. [illegible], Washin[illegible] [illegible], 1958.

Best Available Copy